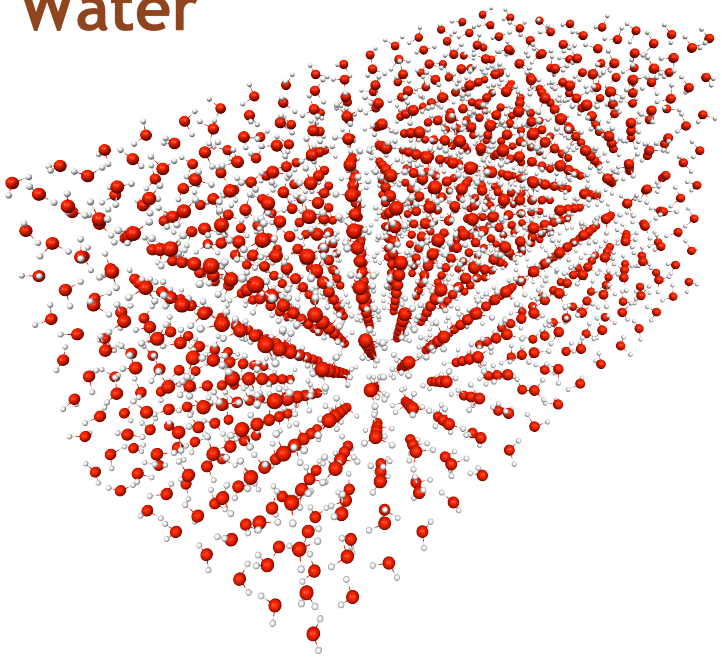
ALCF Early Science Program

Argonne NATIONAL LABORATORY

ESP Kick-Off Workshop
Project Plan Presentation

High Accuracy Predictions of the Bulk Properties of Water



PI: Mark Gordon

Presenter: Graham Fletcher

October 18-19, 2010

Project Overview

Calculate 10 bulk properties of liquid water with *ab initio* cluster simulations

- **Settle scientific debate whether chain/ring networks are a factor in the bulk properties of water**
  - Mira will allow first principles simulations of several thousand water molecules. Intrepid allows simulations of at most 512 waters.
- **Reduce errors in predictions of bulk properties by factor of ten**

Scientific Field: Chemistry

Codes: GAMESS

Project Overview

Calculate 10 bulk properties of liquid water with *ab initio* cluster simulations

- **Settle scientific debate whether chain/ring networks are a factor in the bulk properties of water**
  - Mira will allow first principles simulations of several thousand water molecules. Intrepid allows simulations of at most 512 waters.
- **Reduce errors in predictions of bulk properties by factor of ten**

Bulk Properties targeted:
- Structure
- Density
- Refractive index
- Diffusion coefficient
- Free energy
- Heat capacity
- Dielectric constant
- Vaporization enthalpy
- Isothermal compressibility
- Thermal expansion coefficients

Scientific Field: Chemistry

Codes: GAMESS

Computational Approach, Numerical Methods

- **Solutions to the Schroedinger Equation in a finite atomic basis set**

Computational Approach, Numerical Methods

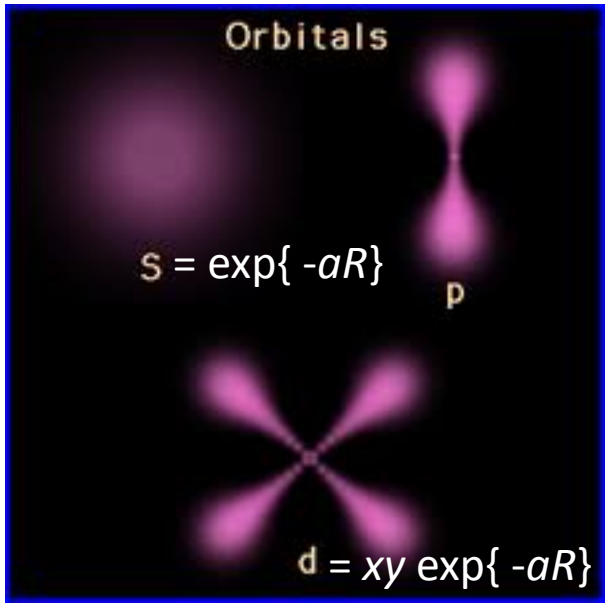
- **Solutions to the Schroedinger Equation in a finite atomic basis set**
  - **Fix atom positions in space while we solve the electronic structure problem**

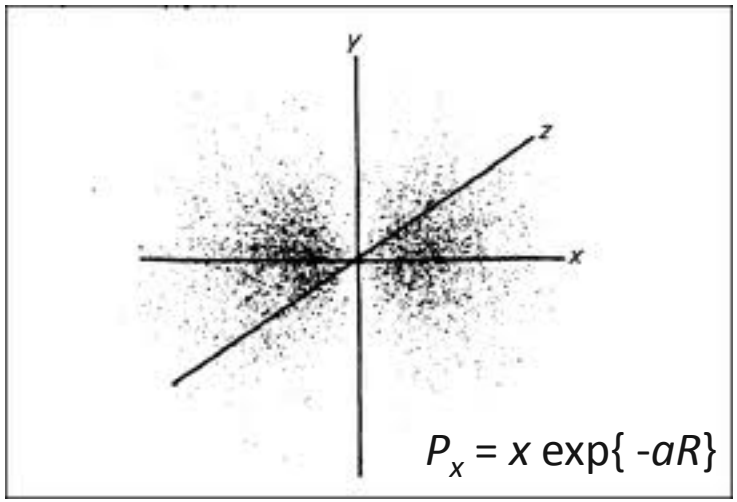
Computational Approach, Numerical Methods

- **Solutions to the Schroedinger Equation in a finite atomic basis set**
  - **Fix atom positions in space while we solve the electronic structure problem**
  - **Model the electrons using functions that obey the rules of quantum mechanics**

Computational Approach, Numerical Methods

- **Solutions to the Schroedinger Equation in a finite atomic basis set**
  - **Fix atom positions in space while we solve the electronic structure problem**
  - **Model the electrons using functions that obey the rules of quantum mechanics**





Computational Approach, Numerical Methods

$\hat{H}\Psi = E\Psi \ldots \langle\Psi|\hat{H}|\Psi\rangle = E\langle\Psi|\Psi\rangle \ldots$

Computational Approach, Numerical Methods

$\hat{H}\Psi = E\Psi \ldots \langle\Psi|\hat{H}|\Psi\rangle = E\langle\Psi|\Psi\rangle \ldots$

$\langle\phi_1\phi_2\ldots|\ldots 1/R_{12}\ldots|\phi_1\phi_2\ldots\rangle \ldots$

Computational Approach, Numerical Methods

$\hat{H}\Psi = E\Psi \ldots \langle\Psi|\hat{H}|\Psi\rangle = E\langle\Psi|\Psi\rangle \ldots$

$\langle\phi_1\phi_2\ldots|\ldots 1/R_{12}\ldots|\phi_1\phi_2\ldots\rangle \ldots$

$\phi_i = \Sigma c_{\upsilon i}\eta_\upsilon \ldots \langle\eta_\upsilon\eta_\nu|1/R_{12}|\eta_\lambda\eta_\sigma\rangle$

Computational Approach, Numerical Methods

$$\hat{H}\Psi = E\Psi \ldots \underline{\langle \Psi | \hat{H} | \Psi \rangle = E \langle \Psi | \Psi \rangle} \ldots$$

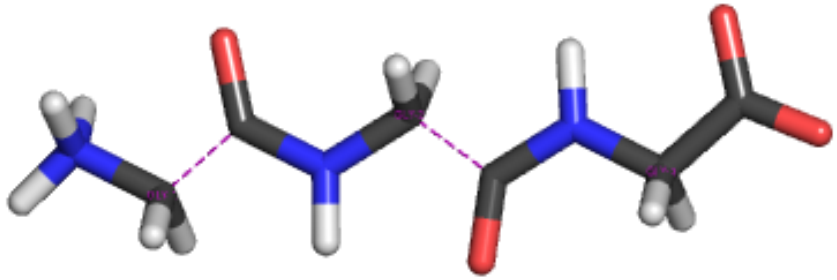
$$\langle \phi_1 \phi_2 \ldots | \ldots 1/R_{12} \ldots | \phi_1 \phi_2 \ldots \rangle \ldots$$

$$\phi_i = \Sigma c_{\upsilon i} \eta_\upsilon \ldots \langle \eta_\upsilon \eta_\nu | 1/R_{12} | \eta_\lambda \eta_\sigma \rangle$$

- **Kernels**
  - Fourth-order: four-index integral computation from analytical formulae
  - Third-order: linear algebra (diagonalization, etc)
  - Higher order ('Electron correlation' models, e.q. MP2):
    - matrix transformation of integral set (fifth-order)

Computational Approach, Numerical Methods

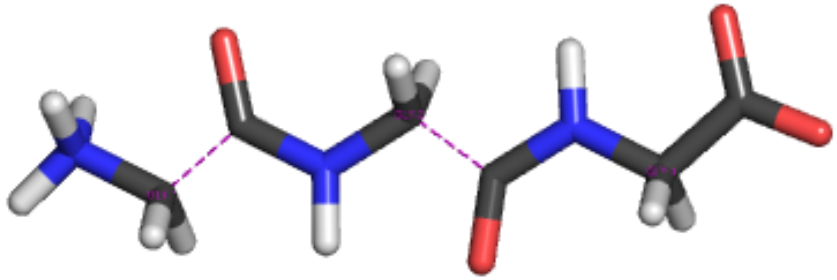
- **Fragment Molecular Orbital (FMO) method**



$$E = \sum_{I}^{N} E_I + \sum_{I>J}^{N} (E_{IJ} - E_I - E_J)$$
$$+ \sum_{I>J>K}^{N} \{(E_{IJK} - E_I - E_J - E_K) - (E_{IJ} - E_I - E_J)$$
$$- (E_{JK} - E_J - E_K) - (E_{KI} - E_K - E_I)\} + \cdots,$$

Computational Approach, Numerical Methods

- **Fragment Molecular Orbital (FMO) method**

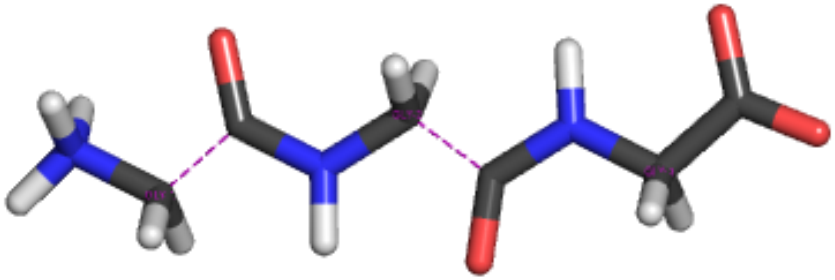


FMO1

$$E = \sum_{I}^{N} E_I + \sum_{I>J}^{N} (E_{IJ} - E_I - E_J)$$
$$+ \sum_{I>J>K}^{N} \{(E_{IJK} - E_I - E_J - E_K) - (E_{IJ} - E_I - E_J)$$
$$- (E_{JK} - E_J - E_K) - (E_{KI} - E_K - E_I)\} + \cdots,$$

Computational Approach, Numerical Methods

- **Fragment Molecular Orbital (FMO) method**

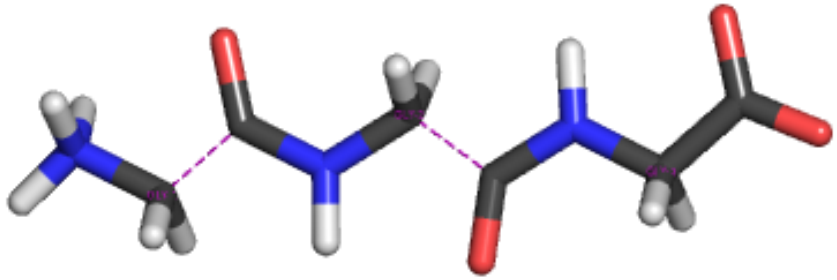


FMO1 FMO2

$$E = \sum_{I}^{N} E_I + \sum_{I>J}^{N} (E_{IJ} - E_I - E_J) + \sum_{I>J>K}^{N} \{(E_{IJK} - E_I - E_J - E_K) - (E_{IJ} - E_I - E_J) - (E_{JK} - E_J - E_K) - (E_{KI} - E_K - E_I)\} + \cdots,$$

Computational Approach, Numerical Methods

- **Fragment Molecular Orbital (FMO) method**



FMO1

FMO2

FMO3

$$E = \sum_{I}^{N} E_I + \sum_{I>J}^{N} (E_{IJ} - E_I - E_J)$$

$$+ \sum_{I>J>K}^{N} \{(E_{IJK} - E_I - E_J - E_K) - (E_{IJ} - E_I - E_J)$$

$$- (E_{JK} - E_J - E_K) - (E_{KI} - E_K - E_I)\} + \cdots,$$

Parallelism and Existing Implementation

- **Coarse and fine-grained parallelism**
  - GAMESS supports general process sub-groups for FMO
  - One-sided data access operations (GET,PUT) support distributed data
  - Static and Dynamic load-balancing are used
- **MPI everywhere (no threads)**
- **I/O (minimal)**
- **Current Performance/Scalability**
  - Scales to 131,072 cores on Intrepid: (FMO2-MP2 forces/aug-cc-pVDZ)

		Racks:	1	2	4	8	16	32
		Cores:	4096	8192	16,384	32,768	65,536	131,072
Waters	Atoms	Basis Functions	Wall-time-to-solution (minutes)					
128	384	5504	8.6	4.8	2.7	1.8		
256	768	11,008	19.8	10.5	5.8	3.4	2.4	
512	1536	22,016		28.9	15.4	8.6	5.1	4.2
1024	3072	44,032			41.1	22.0	12.4	8.0

Library and Tool Dependencies

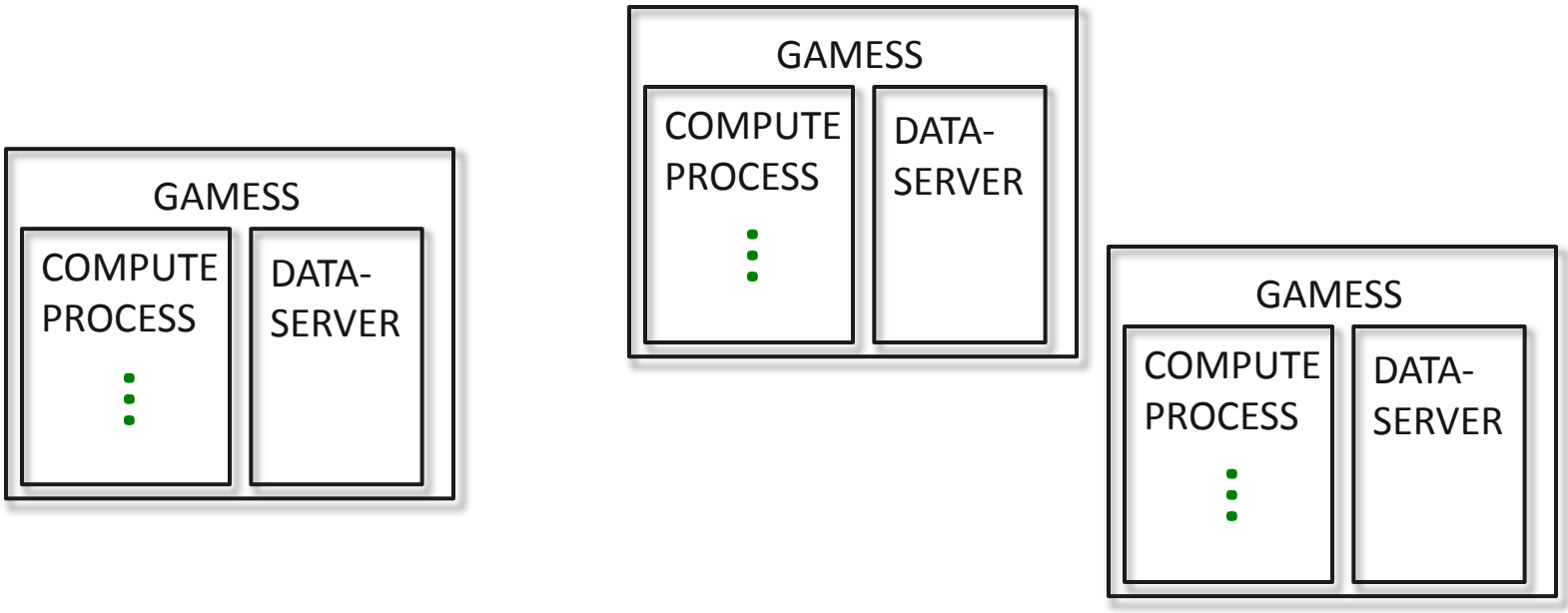
- **Libraries (current)**
  - MPI-1
  - BLAS (ESSL,GOTO)

Anticipated Modifications for Blue Gene/Q

- **Implement thread-based parallelism using OpenMP**
- **Adjustable data-server/node ratio**
- **Non-blocking communications via data-servers**
  - Explicit overlap of communication with computation

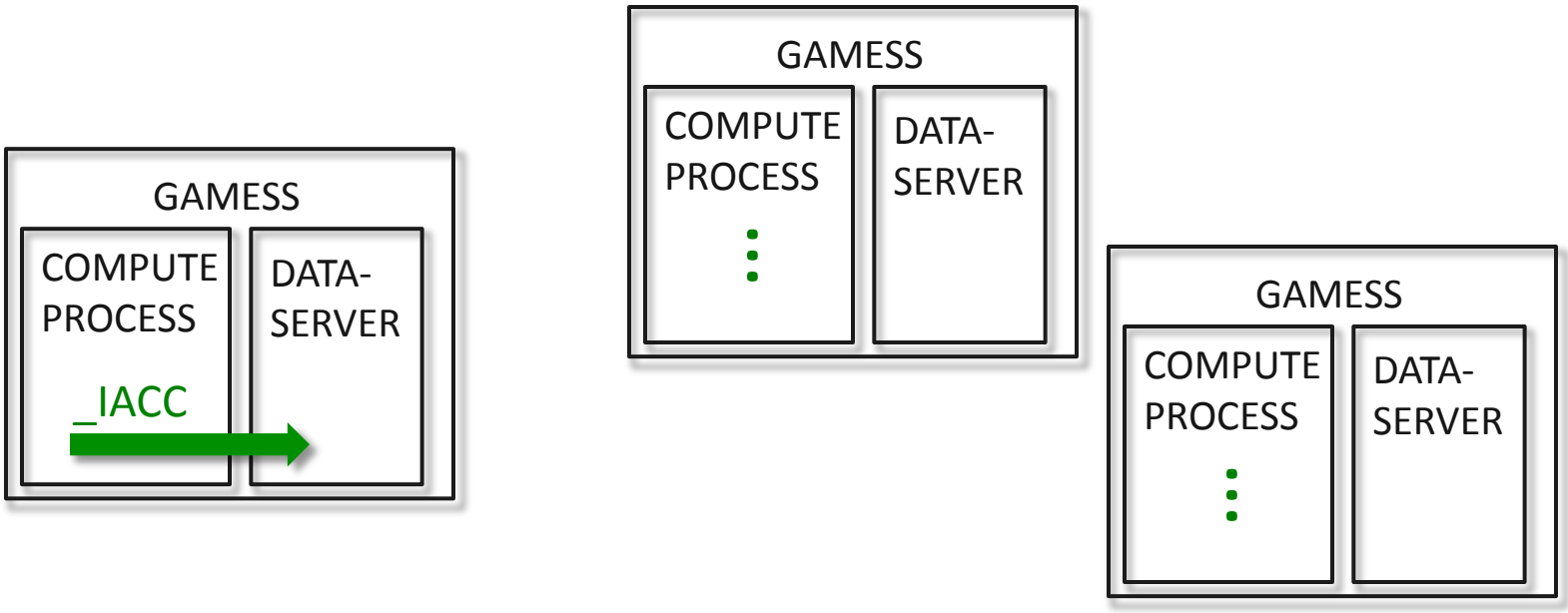
Anticipated Modifications for Blue Gene/Q

- **Implement thread-based parallelism using OpenMP**
- **Adjustable data-server/node ratio**
- **Non-blocking communications via data-servers**
  - Explicit overlap of communication with computation



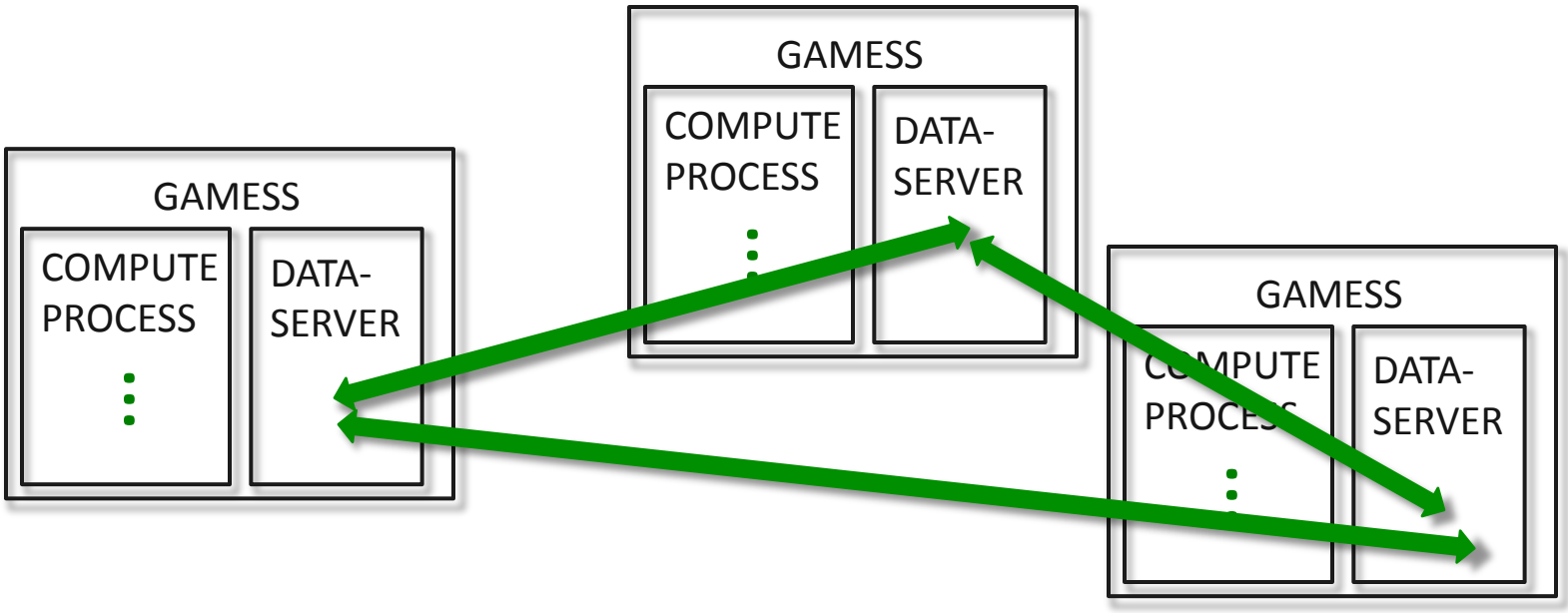
Anticipated Modifications for Blue Gene/Q

- **Implement thread-based parallelism using OpenMP**
- **Adjustable data-server/node ratio**
- **Non-blocking communications via data-servers**
  - Explicit overlap of communication with computation



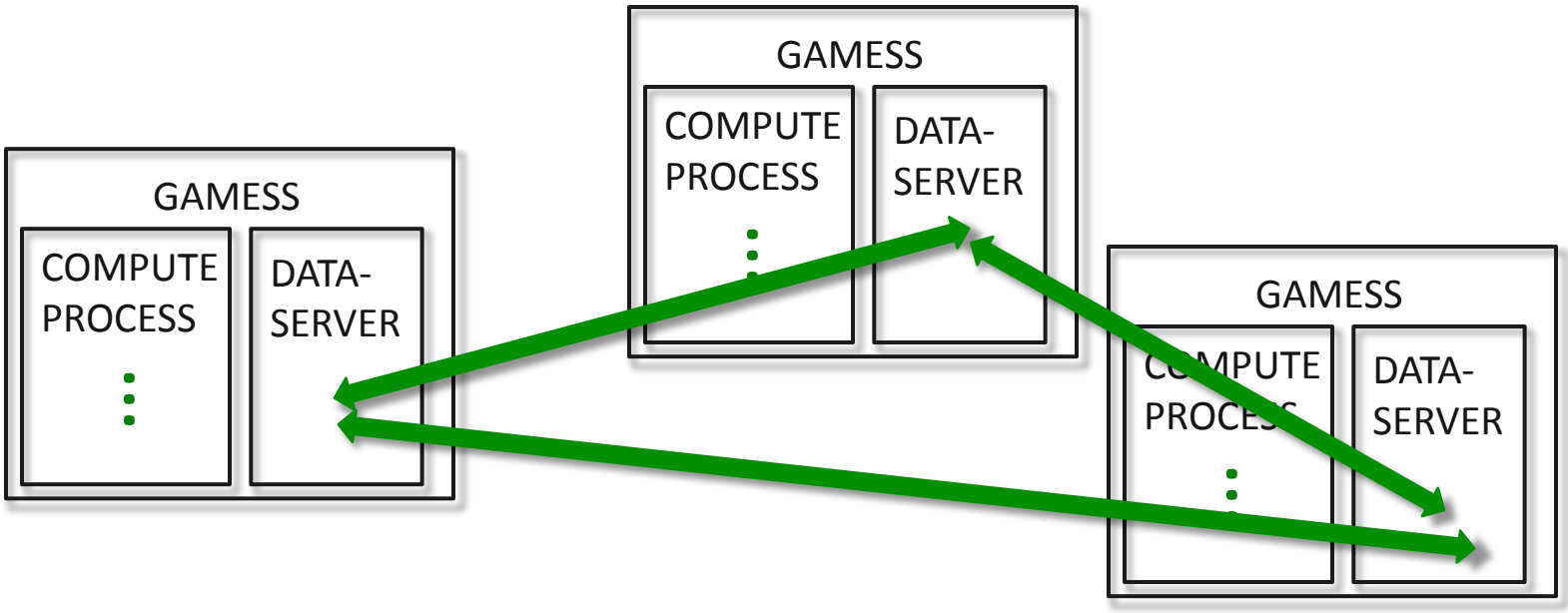
Anticipated Modifications for Blue Gene/Q

- **Implement thread-based parallelism using OpenMP**
- **Adjustable data-server/node ratio**
- **Non-blocking communications via data-servers**
  - Explicit overlap of communication with computation



Anticipated Modifications for Blue Gene/Q

- **Implement thread-based parallelism using OpenMP**
- **Adjustable data-server/node ratio**
- **Non-blocking communications via data-servers**
  - Explicit overlap of communication with computation



- **Performance and scaling needed to run proposed problem on Mira:**
  - <5 wallclock minutes per timestep

Plan for Next 6 Months Effort

- **Help find and hire a project postdoc**
- **Code developments begin on Blue Gene/P**
  - Data-server(DS)/node
  - DS-DS communications
- **Introduce OpenMP in main code kernels:**
  - Integral Computation
  - Construction of 'Fock' matrix
- **Benchmark performance on BG/P w.r.t.:**
  - Threads/MPI task
  - Data-servers/node
- **Use projections to estimate performance on BG/Q**